

Open Research Online

The Open University's repository of research publications and other research outputs

A geometric representation of continued fractions

Journal Item

How to cite:

Beardon, Alan F. and Short, Ian (2014). A geometric representation of continued fractions. *American Mathematical Monthly*, 121(5) pp. 391–402.

For guidance on citations see [FAQs](#).

© 2014 Mathematical Association of America

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.4169/amer.math.monthly.121.05.391>

<http://www.jstor.org/stable/10.4169/amer.math.monthly.121.05.391>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

A Geometric Representation of Continued Fractions

Alan F. Beardon and Ian Short

Abstract. Inspired by work of Ford, we describe a geometric representation of real and complex continued fractions by chains of horocycles and horospheres in hyperbolic space. We explore this representation using the isometric action of the group of Möbius transformations on hyperbolic space, and prove a classical theorem on continued fractions.

1. INTRODUCTION. In this paper, we consider infinite complex continued fractions of the form

$$\mathbf{K}(b_n) = b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \frac{1}{b_4 + \cdots}}},$$

where b_1, b_2, \dots are complex numbers. We say that this is an integer continued fraction if all b_i are integers, and a real continued fraction if all b_i are real. It has long been recognized that we can study real and complex continued fractions from a geometric point of view by using Möbius transformations; here, we show how to represent real continued fractions by chains of horocycles in the hyperbolic plane, and complex continued fractions by chains of horospheres in hyperbolic space. The origin of this idea is in Ford's well-known paper [5] where he used such a representation to study integer continued fractions. Ford constructs the horocycles at rational points in an elementary way; indeed, he says, "Perhaps the author owes an apology to the reader for asking him to lend his attention to so elementary a subject, ...". On the other hand, he also says that his original idea was motivated by Bianchi's study of the Picard group. This material is at a deeper level, and Ford uses this in [4] where he discusses horospheres in three-dimensional hyperbolic space that are based at the Gaussian integers. Here we follow a similar path from an elementary representation of real continued fractions by horocycles to a deeper study of the representation of complex continued fractions by horospheres in three-dimensional hyperbolic space.

2. CONTINUED FRACTIONS AND MÖBIUS MAPS. We begin by describing the relationship between continued fractions and certain sequences of matrices and Möbius maps. Let b_1, b_2, \dots be as above. Moreover, let A_0, A_1, A_2, \dots and B_0, B_1, B_2, \dots , where $A_0 = 1$ and $B_0 = 0$, be given by

$$\begin{pmatrix} A_n & A_{n-1} \\ B_n & B_{n-1} \end{pmatrix} = \begin{pmatrix} b_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix}. \quad (2.1)$$

<http://dx.doi.org/10.4169/amer.math.monthly.121.05.391>
MSC: Primary 40A15, Secondary 30F45, 30B70

Taking determinants gives $|A_n B_{n-1} - A_{n-1} B_n| = 1$. We also define $t_n(z) = b_n + 1/z$ and $T_n(z) = (A_n z + A_{n-1})/(B_n z + B_{n-1})$ for $n = 1, 2, \dots$, and then, corresponding to (2.1), we have

$$T_n = t_1 \circ t_2 \circ \dots \circ t_n.$$

Evaluating this equation at ∞ gives

$$T_n(\infty) = \frac{A_n}{B_n} = b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \dots + \frac{1}{b_n}}}.$$

Clearly, we are working in the extended complex plane, and $\mathbf{K}(b_n)$ converges if the sequence $T_n(\infty)$ converges; otherwise, it *diverges*.

Note that we can recapture the coefficients b_n from the maps T_n . Indeed,

$$b_n = t_n(\infty) = T_{n-1}^{-1} T_n(\infty),$$

and since we know the matrix representations for T_n and T_{n-1} , we can simplify the term on the right to obtain $|A_n B_{n-2} - B_n A_{n-2}| = |b_n|$. This gives

$$|T_n(\infty) - T_{n-2}(\infty)| = \frac{|b_n|}{|B_{n-2}| |B_n|}. \quad (2.2)$$

3. CHAINS OF HOROCYCLES. In this section, we shall assume that the coefficients b_1, b_2, \dots are real numbers (but not necessarily integers). Let \mathbb{H} denote the upper half of the complex plane \mathbb{C} . A *horocycle* in \mathbb{H} refers to either a circle in \mathbb{C} that is tangent to the real axis and otherwise lies in \mathbb{H} , or else a horizontal line in \mathbb{H} (which has constant imaginary part) with the point ∞ attached. The *base point* of the horocycle is the point of tangency in the first case, and ∞ in the latter case. A simple application of Pythagoras' theorem gives the following lemma.

Lemma 3.1. *Two horocycles with distinct real base points x and y , and Euclidean radii r and s , are tangent if and only if*

$$|x - y|^2 = 4rs.$$

Next, using the matrix entries A_n and B_n of (2.1), we establish a correspondence between continued fractions and chains of horocycles in the hyperbolic plane. Let Π_0 denote the horocycle $\{z : \text{Im}[z] = 1\} \cup \{\infty\}$. Given a continued fraction $\mathbf{K}(b_n)$, define Π_n , for each positive integer n , to be the horocycle with base point A_n/B_n and Euclidean radius $1/(2B_n^2)$ provided $B_n \neq 0$, and if $B_n = 0$, then define $\Pi_n = \{z : \text{Im}[z] = A_n^2\} \cup \{\infty\}$. Suppose that $B_n, B_{n-1} \neq 0$; then

$$\left| \frac{A_n}{B_n} - \frac{A_{n-1}}{B_{n-1}} \right|^2 = \frac{|A_n B_{n-1} - A_{n-1} B_n|^2}{B_n^2 B_{n-1}^2} = 4 \left(\frac{1}{2B_n^2} \right) \left(\frac{1}{2B_{n-1}^2} \right), \quad (3.1)$$

and so Lemma 3.1 implies that Π_n and Π_{n-1} are tangent. Likewise, we can check that Π_n and Π_{n-1} are tangent if one of B_n or B_{n-1} is 0 (they cannot both be 0). We say that a sequence of horocycles Π_0, Π_1, \dots of this type (where Π_n and Π_{n-1} are tangent and Π_0 is given by $\text{Im}[z] = 1$) is a *chain of horocycles*. The first five horocycles in a typical chain are shown in Figure 3.1. Notice that nonconsecutive horocycles may overlap.

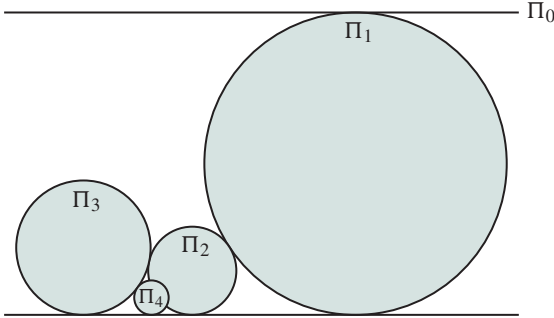


Figure 3.1. The beginnings of a chain of horocycles

We have ignored the algebraic details of the special case $B_n = 0$, which occurs precisely when $T_n(\infty) = \infty$. We return to this issue afresh in Section 6 with a geometric approach that avoids the need to distinguish the point ∞ . In Section 4, we study continued fractions with positive coefficients b_n , in which case the denominators B_n are also positive, and all horocycles Π_n are Euclidean circles.

Each horocycle in a chain corresponding to an integer continued fraction is a *Ford circle*; that is, its base point is a reduced rational p/q , and its Euclidean radius is $1/(2q^2)$. Ford introduced chains of Ford circles in [5]. Ford circles never overlap. A collection of Ford circles is shown in Figure 3.2, and the first few horocycles in a chain are shaded in a darker color.

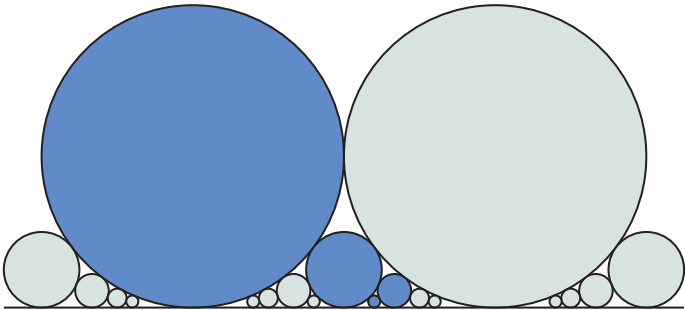


Figure 3.2. The beginnings of a chain of Ford circles

4. A CONVERGENCE THEOREM. The following well-known theorem (originally due to Seidel and Stern) is found in many classic texts on continued fractions, such as [6, Theorem 166], [8, Theorem 10], and [9, Theorem 3.13].

Theorem 4.1. *Suppose that b_1, b_2, \dots are positive numbers. If $\sum b_n$ diverges, then the continued fraction $\mathbf{K}(b_n)$ converges.*

This result shows that continued fractions with positive integer coefficients converge. In this section we illustrate and prove the theorem using chains of horocycles.

Each transformation $t_n(z) = b_n + 1/z$, where now $b_n > 0$, maps the interval $[0, \infty]$ within itself. It follows that T_n also maps $[0, \infty]$ within itself, and so the point $T_n(b_{n+1})$ lies between $T_n(0)$ and $T_n(\infty)$. Now

$$T_n(b_{n+1}) = T_{n+1}(\infty) \quad \text{and} \quad T_n(0) = T_{n-1}(\infty),$$

and therefore we have shown that $T_{n+1}(\infty)$ lies between $T_{n-1}(\infty)$ and $T_n(\infty)$. Since trivially $T_1(\infty) < T_2(\infty)$, we deduce that

$$0 < T_1(\infty) < T_3(\infty) < \cdots < T_{2n-1}(\infty) < T_{2n}(\infty) < \cdots < T_4(\infty) < T_2(\infty). \quad (4.1)$$

This implies that there are real numbers α and β , with $\alpha \leq \beta$, such that

$$T_{2n-1}(\infty) \rightarrow \alpha \quad \text{and} \quad T_{2n}(\infty) \rightarrow \beta \quad (4.2)$$

as $n \rightarrow \infty$, and $\mathbf{K}(b_n)$ converges if and only if $\alpha = \beta$.

Let us now examine the chain of horocycles Π_0, Π_1, \dots for continued fractions with positive coefficients. By (3.1), the Euclidean radii r_n and r_{n-1} of the horocycles Π_n and Π_{n-1} satisfy

$$|T_n(\infty) - T_{n-1}(\infty)|^2 = 4r_n r_{n-1}. \quad (4.3)$$

Therefore,

$$4r_n r_{n-1} = |T_n(\infty) - T_{n-1}(\infty)|^2 < |T_{n-1}(\infty) - T_{n-2}(\infty)|^2 = 4r_{n-1} r_{n-2},$$

and so $r_n < r_{n-2}$. We deduce that both sequences r_1, r_3, \dots and r_2, r_4, \dots are decreasing (see Figure 4.1).

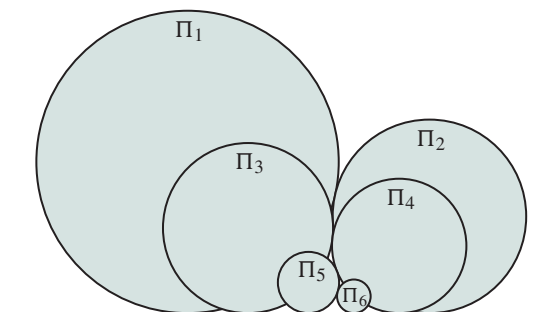


Figure 4.1. The beginnings of a chain of horocycles when all b_n are positive

Suppose that $\mathbf{K}(b_n)$ diverges; then the limits α and β from (4.2) are distinct. To prove Theorem 4.1, we must show that $\sum b_n$ converges. By (4.3),

$$4r_n r_{n-1} = |T_n(\infty) - T_{n-1}(\infty)|^2 > |\alpha - \beta|^2.$$

This implies that the decreasing sequences r_1, r_3, \dots and r_2, r_4, \dots both converge to positive constants. Now, from (2.2), we see that

$$|T_n(\infty) - T_{n-2}(\infty)| = 2b_n \sqrt{r_n r_{n-2}}.$$

We know from (4.1) that $\sum |T_{n-2}(\infty) - T_n(\infty)|$ converges, and since the sequence r_1, r_2, \dots is bounded below by a positive constant, we deduce that $\sum b_n$ also converges. This completes the proof of Theorem 4.1.

5. MÖBIUS TRANSFORMATIONS AND HYPERBOLIC GEOMETRY. The relationship between a continued fraction and its chain of horocycles can be better understood by using the action of complex Möbius transformations on three-dimensional hyperbolic space. We briefly review some of the theory of Möbius transformations and hyperbolic geometry, which can be found in more detail in [1].

The set

$$\mathbb{H}^3 = \{(x, y, t) \in \mathbb{R}^3 : t > 0\}$$

is a model of three-dimensional hyperbolic space when equipped with the hyperbolic metric $\sqrt{dx^2 + dy^2 + dt^2}/t$. The corresponding distance function ϱ on \mathbb{H}^3 is given by

$$\varrho(u, v) = \inf_{\gamma} \int_{\gamma} \frac{\sqrt{dx^2 + dy^2 + dt^2}}{t},$$

where the infimum is taken over all smooth paths γ from u to v in \mathbb{H}^3 . Unlike the Euclidean metric on \mathbb{H}^3 , the hyperbolic metric is complete. Let us identify the point $(x, y, 0)$ with the complex number $z = x + iy$. The complex plane \mathbb{C} is then identified with the Euclidean plane $t = 0$. The group of Möbius transformations acts on $\mathbb{C} \cup \{\infty\}$, and we now describe how this action can be extended to \mathbb{H}^3 .

Consider a Möbius transformation $f(z) = (az + b)/(cz + d)$, where a, b, c , and d are complex numbers with $ad - bc = 1$. Define $j = (0, 0, 1)$. Then, (x, y, t) can be represented by the quaternion $z + tj$, and the action of f on \mathbb{H}^3 is given by

$$f(z + tj) = \frac{(az + b)(\overline{cz + d}) + a\bar{c}t^2 + tj}{|cz + d|^2 + |c|^2t^2}; \quad (5.1)$$

see [1, Section 4.1]. This action *preserves the hyperbolic metric on \mathbb{H}^3* . In other words,

$$\varrho(f(u), f(v)) = \varrho(u, v)$$

for any pair of points u and v in \mathbb{H}^3 . In fact, every conformal isometry of \mathbb{H}^3 arises in this way.

We end this section by examining the actions on \mathbb{H}^3 of two specific types of Möbius transformation. First, consider a translation $f(z) = z + b$. Then

$$f(z + tj) = z + b + tj.$$

That is, f also acts as a translation by b on \mathbb{H}^3 . Next, suppose that $f(z) = 1/z$. We must write this map as $f(z) = i/(iz)$ in order to satisfy the condition $ad - bc = 1$. Then

$$f(z + tj) = \frac{\bar{z} + tj}{|z|^2 + t^2}. \quad (5.2)$$

Geometrically, f is an inversion in the unit sphere (which preserves \mathbb{H}^3) followed by reflection in the Euclidean plane $y = 0$ (which also preserves \mathbb{H}^3).

6. CHAINS OF HOROSPHERES. With the geometry developed in the previous section, we are now able to take a more sophisticated view of chains of horocycles associated to continued fractions. This time we assume that the coefficients b_n of our continued fraction are arbitrary *complex* numbers. A *horosphere* in \mathbb{H}^3 is either a Euclidean sphere in \mathbb{R}^3 that is tangent to \mathbb{C} and otherwise lies in \mathbb{H}^3 , or else a Euclidean plane in \mathbb{H}^3 , parallel to \mathbb{C} , with the point ∞ attached. The *base point* of the horosphere is the point of tangency in the first case, and ∞ in the latter case. The image of a horosphere under a Möbius transformation is another horosphere.

Let

$$\Sigma_0 = \{z + j : z \in \mathbb{C}\} \cup \{\infty\}.$$

We can calculate the image of Σ_0 under a Möbius transformation using the representation by quaternions described in the previous section. Let

$$\text{ht}(z + tj) = t,$$

so that ht measures the ‘height’ of a point $z + tj$ in \mathbb{H}^3 .

Lemma 6.1. *Suppose that $f(z) = (az + b)/(cz + d)$, where $ad - bc = 1$. If $c \neq 0$, then $f(\Sigma_0)$ has base point a/c and Euclidean radius $1/(2|c|^2)$. If $c = \infty$, then $f(\Sigma_0) = \{z + |a|^2j : z \in \mathbb{C}\} \cup \{\infty\}$.*

Proof. The base point of $f(\Sigma_0)$ is $f(\infty)$. Suppose first that $c \neq 0$. Then $f(\infty) = a/c$, so $f(\Sigma_0)$ is a Euclidean sphere. By (5.1), we have

$$\text{ht}(f(z + j)) = \frac{1}{|cz + d|^2 + |c|^2}.$$

The maximum value of this expression is $1/|c|^2$ (when $z = -d/c$), and hence $f(\Sigma_0)$ has Euclidean radius $1/(2|c|^2)$.

Suppose now that $c = 0$. Then $ad = 1$, so $d = 1/a$. From (5.1) we obtain

$$f(z + j) = a^2z + ab + |a|^2j.$$

It follows that $f(\Sigma_0) = \{z + |a|^2j : z \in \mathbb{C}\} \cup \{\infty\}$. ■

Let us apply Lemma 6.1 to the map $t_n(z) = b_n + 1/z$. First we must write t_n in the form $t_n(z) = (ib_nz + i)/(iz + 0)$, to satisfy the condition $ad - bc = 1$. Then Lemma 6.1 says that $t_n(\Sigma_0)$ has base point b_n and Euclidean radius $1/2$. Hence $t_n(\Sigma_0)$ is tangent to Σ_0 at the point $t_n(j) = b_n + j$, as illustrated in Figure 6.1.

Recall that $T_n = t_1 \circ t_2 \circ \cdots \circ t_n$. The horospheres Σ_0 and $t_n(\Sigma_0)$ are tangent at the point $t_n(j)$, which implies that the horospheres $T_{n-1}(\Sigma_0)$ and $T_n(\Sigma_0) = T_{n-1}(t_n(\Sigma_0))$ are tangent at the point $T_n(j)$. A *chain of horospheres* is a sequence of horospheres

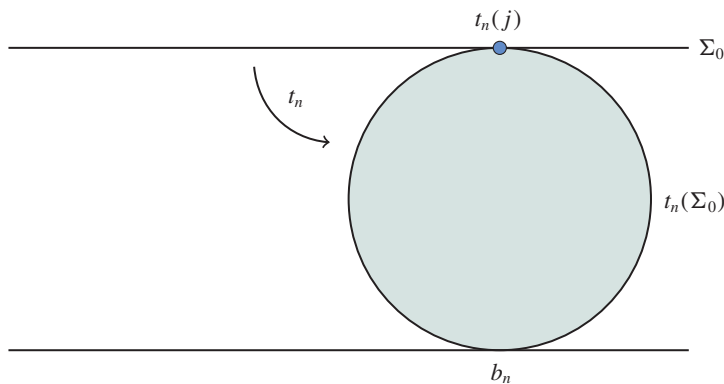


Figure 6.1. The horospheres Σ_0 and $t_n(\Sigma_0)$ are tangent at the point $t_n(j)$.

$\Sigma_0, \Sigma_1, \Sigma_2, \dots$, where Σ_n is tangent to Σ_{n-1} for $n = 1, 2, \dots$. We have proven that $T_n(\Sigma_0)$ is a chain of horospheres (see Figure 6.2).

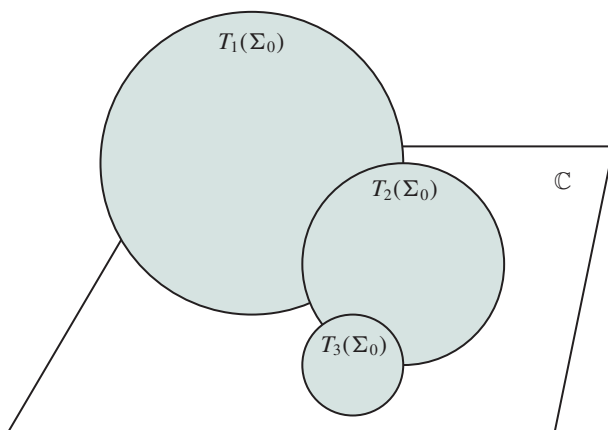


Figure 6.2. The beginnings of a chain of horospheres

The base points z_0, z_1, z_2, \dots of a chain of horospheres satisfy $z_0 = \infty$ and $z_n \neq z_{n-1}$ for $n = 1, 2, \dots$. Conversely, any sequence of points of this type determines a unique chain of horospheres $\Sigma_0, \Sigma_1, \dots$ (because Σ_0 is fixed and Σ_n is determined inductively by the conditions that it has base point z_n and is tangent to Σ_{n-1}). We use this observation to establish a correspondence between complex continued fractions and chains of horospheres.

Theorem 6.2. *Given a continued fraction $\mathbf{K}(b_n)$ with complex coefficients, the sequence $\Sigma_0, T_1(\Sigma_0), T_2(\Sigma_0), \dots$ is a chain of horospheres. Conversely, given a chain of horospheres $\Sigma_0, \Sigma_1, \Sigma_2, \dots$ there is a unique continued fraction $\mathbf{K}(b_n)$ with $T_n(\Sigma_0) = \Sigma_n$ for $n = 1, 2, \dots$*

Proof. We have only to prove the converse statement. Suppose that $\Sigma_0, \Sigma_1, \Sigma_2, \dots$ is a chain of horospheres with base points z_0, z_1, z_2, \dots . Define the sequence b_n and corresponding maps $t_n(z) = b_n + 1/z$ and $T_n = t_1 \circ t_2 \circ \dots \circ t_n$ inductively by $b_1 = z_1$

and $b_n = T_{n-1}^{-1}(z_n)$. Then $t_n(\infty) = b_n = T_{n-1}^{-1}(z_n)$, and so $T_n(\infty) = z_n$. The condition $z_n \neq z_{n-1}$ ensures that $b_n \neq \infty$, because $b_n = T_{n-1}^{-1}(z_n) \neq T_{n-1}^{-1}(z_{n-1}) = \infty$. Now, $\Sigma_0, T_1(\Sigma_0), T_2(\Sigma_0), \dots$ is a chain of horospheres with base points z_0, z_1, z_2, \dots , and it follows from the remark before this theorem that $T_n(\Sigma_0) = \Sigma_n$. The uniqueness of $\mathbf{K}(b_n)$ is a consequence of the inductive equations $b_1 = z_1$ and $b_n = T_{n-1}^{-1}(z_n)$. ■

Using Lemma 6.1, we can describe the horosphere $T_n(\Sigma_0)$ in terms of the coefficients of T_n . We recall the sequences of complex numbers A_0, A_1, \dots and B_0, B_1, \dots determined by the matrix equation (2.1).

Lemma 6.3. *If $B_n \neq 0$, then $T_n(\Sigma_0)$ has base point A_n/B_n and Euclidean radius $1/(2|B_n|^2)$. If $B_n = 0$, then $T_n(\Sigma) = \{z + |A_n|^2 j : z \in \mathbb{C}\} \cup \{\infty\}$.*

Proof. This follows immediately from Lemma 6.1, because

$$T_n(z) = (A_n z + A_{n-1}) / (B_n z + B_{n-1}) \quad \text{and} \quad |A_n B_{n-1} - A_{n-1} B_n| = 1. \quad \blacksquare$$

The horospheres $T_n(\Sigma_0)$ and $T_{n-1}(\Sigma_0)$ are tangent at the point $T_n(j)$. The base points of these two horospheres are $T_n(\infty)$ and $T_{n-1}(\infty) = T_n(0)$. The hyperbolic line between ∞ and 0 in \mathbb{H}^3 (a Euclidean half-line), which contains j , is mapped by T_n to the hyperbolic line between $T_n(\infty)$ and $T_n(0)$ in \mathbb{H}^3 ; see Figure 6.3.

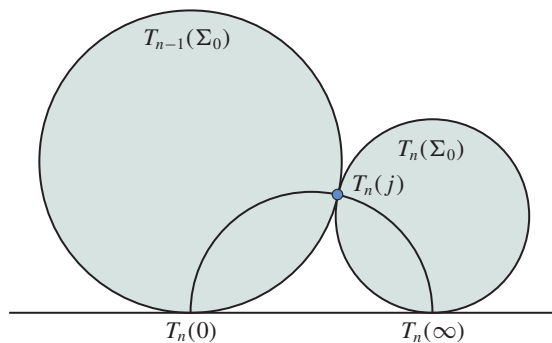


Figure 6.3. The horospheres $T_{n-1}(\Sigma_0)$ and $T_n(\Sigma_0)$ are tangent at the point $T_n(j)$.

If the continued fraction $\mathbf{K}(b_n)$ converges to a complex number p , then $T_n(\infty) \rightarrow p$ and $T_n(0) \rightarrow p$ as $n \rightarrow \infty$. Because $T_n(j)$ lies on the hyperbolic line between $T_n(\infty)$ and $T_n(0)$, we see that $T_n(j)$ approaches the boundary of hyperbolic space as $n \rightarrow \infty$. In particular, $\varrho(j, T_n(j)) \rightarrow \infty$ as $n \rightarrow \infty$. (In fact, $T_n(j)$ converges to p in the Euclidean metric.) If $p = \infty$, then it is still true that $\varrho(j, T_n(j)) \rightarrow \infty$ as $n \rightarrow \infty$, but to prove this rigorously, we should really use the ball model of three-dimensional hyperbolic space with the spherical metric, because in this model the point ∞ no longer has any special geometric significance.

7. THE CONVERSE TO THEOREM 4.1. Here we prove the converse to Theorem 4.1, namely that if $\mathbf{K}(b_n)$ converges, then $\sum b_n$ diverges. In fact, we prove a stronger result, known as the Stern–Stolz theorem (see [9, Theorem 3.1]), which

says that given any sequence of *complex* numbers b_1, b_2, \dots , if $\mathbf{K}(b_n)$ converges, then $\sum |b_n|$ diverges. Our proof is essentially the same as that in [2].

Observe that $t_n(z) = f(b_n + z)$, where $f(z) = 1/z$, and f fixes the point j . Let Γ denote the collection of all smooth paths in \mathbb{H}^3 from j to $b_n + j$, and let δ denote the Euclidean line segment between j and $b_n + j$. Thus, $\delta \in \Gamma$. Let $\zeta = z + tj$; then, because f is a hyperbolic isometry, we have

$$\varrho(j, t_n(j)) = \varrho(f(j), f t_n(j)) = \varrho(j, b_n + j) = \inf_{\gamma \in \Gamma} \int_{\gamma} \frac{|d\zeta|}{t} \leq \int_{\delta} \frac{|d\zeta|}{t} = |b_n|.$$

Now, T_{n-1} is also a hyperbolic isometry, so $\varrho(j, t_n(j)) = \varrho(T_{n-1}(j), T_n(j))$. Thus, using the triangle inequality, we find that

$$\begin{aligned} \rho(j, T_n(j)) &\leq \rho(j, T_1(j)) + \rho(T_1(j), T_2(j)) + \dots + \rho(T_{n-1}(j), T_n(j)) \\ &\leq |b_1| + |b_2| + \dots + |b_n|. \end{aligned}$$

But $\mathbf{K}(b_n)$ converges, and we saw at the end of Section 6 that this implies that $\rho(j, T_n(j)) \rightarrow \infty$ as $n \rightarrow \infty$. Therefore, $\sum |b_n|$ diverges.

8. THE VERTICAL HYPERBOLIC PLANE. We have seen how to represent a real continued fraction by a chain of horocycles in the hyperbolic plane, and a complex continued fraction by a chain of horospheres in three-dimensional hyperbolic space. However, there is a significant difference between these two constructions. In the real case, the horocycles are constructed in an ad hoc manner; in the complex case, the horospheres are constructed by using the isometric action of the Möbius group on hyperbolic space. We should ask, therefore, whether in the real case we can obtain the chain of horocycles by a group action in the hyperbolic plane. Now, there are many papers on integer continued fractions that refer to the action of the modular group (of Möbius maps $z \mapsto (az + b)/(cz + d)$, where a, b, c , and d are integers with $ad - bc = 1$) on the upper half-plane \mathbb{H} , namely $\{x + iy : y > 0\}$, which is regarded as the hyperbolic plane with metric $|dz|/y$. However, even this is not entirely satisfactory, because when studying continued fractions in this way, we are inevitably led to consider maps of the form $z \mapsto b + 1/z$, and such maps do not leave \mathbb{H} invariant (they interchange the upper and lower half-planes). Some papers also use the extended modular group (in which we only require the integers a, b, c and d to satisfy $ad - bc = \pm 1$), and the same problem arises for this group.

The real problem here is that the maps $z \mapsto b + 1/z$ are strictly loxodromic, and *there is no disc or half-plane in the extended complex plane that is invariant under a strictly loxodromic map*; see [1, Theorem 4.3.4]. It follows that if we wish to describe real continued fractions purely in terms of the group action of isometries of the hyperbolic plane, then we have to look for this plane elsewhere. In fact, there is such a plane and it lies in three-dimensional hyperbolic space \mathbb{H}^3 . Let

$$\mathbb{H}^\perp = \{x + tj : x \in \mathbb{R}, t > 0\};$$

we call this the *vertical hyperbolic plane*, and it is the set of those points $z + tj$ in \mathbb{H}^3 with $y = 0$. It is well known that a Möbius map f leaves the extended real line invariant if and only if it can be written in the form $f(z) = (az + b)/(cz + d)$, where the coefficients a, b, c , and d are real, and $ad - bc = \pm 1$. Equally, this is so if and only if f leaves \mathbb{H}^\perp invariant. It follows that when considering real continued fractions (or,

more generally, real Möbius maps) it might be advantageous to consider their action on the *invariant* plane \mathbb{H}^\perp . Indeed, \mathbb{H}^\perp , with the hyperbolic metric induced from \mathbb{H}^3 , is a hyperbolic plane in \mathbb{H}^3 , and the group of Möbius maps with real coefficients acts faithfully as a group of hyperbolic isometries on \mathbb{H}^\perp . Let us examine, for example, the action of $g(z) = 1/\bar{z}$ (which does not preserve the upper half-plane \mathbb{H}) on \mathbb{H}^\perp . By (5.2), we have

$$g(x + tj) = \frac{x + tj}{x^2 + t^2},$$

so g acts on \mathbb{H}^\perp as inversion across the unit circle $x^2 + t^2 = 1$ (and this action on \mathbb{H}^\perp is anticonformal). More generally, the map f above is conformal on \mathbb{H}^\perp if $ad - bc = 1$, and is anticonformal if $ad - bc = -1$.

This view also throws light on the use of the extended modular group in continued fraction theory. The extended modular group is an extension of the modular group, but in discrete group theory, it is the Picard group (the set of Möbius maps with a, b, c , and d Gaussian integers, and $ad - bc = 1$), which acts on three-dimensional hyperbolic space, that is regarded as the natural analogue of the modular group. To complete the link between these three groups, we only have to note that the extended modular group is simply the subgroup of the Picard group that leaves the extended real line invariant. Thus, from a different perspective we might consider replacing the action of the extended modular group on the complex plane by its action (as a subgroup of the Picard group) on the vertical hyperbolic plane.

Let us now return to real continued fractions. Despite the discussion given in Section 3, we can regard the real coefficients b_n as complex coefficients, and construct horospheres and so on, as in Section 6. If we do this, and then intersect the resulting horospheres with the vertical hyperbolic plane \mathbb{H}^\perp , we will obtain exactly the same result as we achieved in Section 3, except that now the horocycles in \mathbb{H}^\perp are given by a group action. In this way, we see that the geometric theory of real continued fractions really is a special case of the geometric theory of complex continued fractions, for both are described by the same group action on the same space.

When the coefficients b_n are real, the sequence $T_1(\infty), T_2(\infty), \dots$ lies in the extended real line \mathbb{R}_∞ , and the horosphere $T_n(\Sigma_0)$ intersects \mathbb{H}^\perp orthogonally in a circle Π'_n , which is a horocycle in the vertical hyperbolic plane (see Figure 8.1). Let Π'_0 denote the horocycle given by $t = 1$ in \mathbb{H}^\perp . Then Π'_n and Π'_{n-1} are tangent for $n = 1, 2, \dots$, and so Π'_0, Π'_1, \dots is a chain of horocycles. The base point of Π'_n is $T_n(\infty) = A_n/B_n$ and, by Lemma 6.3, the Euclidean radius of Π'_n is $1/(2B_n^2)$.

Essentially, we have recovered the original chain of horocycles associated to a real continued fraction described in Section 3, except now the horocycles lie in \mathbb{H}^\perp rather than \mathbb{H} , and using the isometric action of Möbius transformations on hyperbolic space, we have a much deeper understanding of the geometry.

9. CONCLUDING REMARKS. The representation of continued fractions by chains of horospheres may shed light on other results in continued fraction theory. As a basic example, suppose that the horospheres $\Sigma_0, T_1(\Sigma_0), T_2(\Sigma_0), \dots$ corresponding to a continued fraction $\mathbf{K}(b_n)$ are all Euclidean spheres, and

$$\sum_{n=1}^{\infty} \text{rad}[T_n(\Sigma_0)] < +\infty,$$

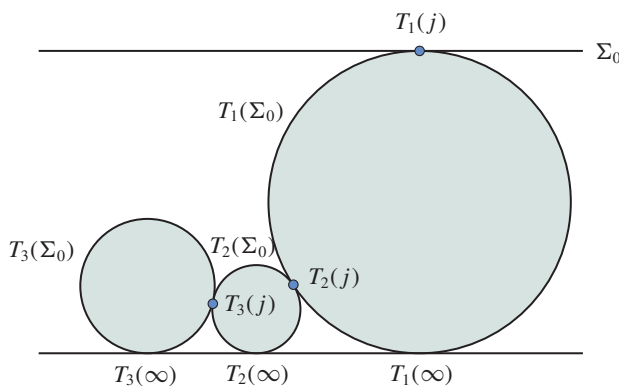


Figure 8.1. The beginnings of a chain of horospheres when all b_n are real

where $\text{rad}[T_n(\Sigma_0)]$ is the Euclidean radius of $T_n(\Sigma_0)$. Then, geometrically, it is clear that $\mathbf{K}(b_n)$ converges. Lemma 6.3 tells us that $\text{rad}[T_n(\Sigma_0)] = 1/(2|B_n|^2)$, and thus we recover the simple result that convergence of $\sum 1/|B_n|^2$ implies convergence of $\mathbf{K}(b_n)$.

Another advantage of the geometric approach (using the horospheres $T_n(\Sigma_0)$) rather than the algebraic approach (using the coefficients B_n) is that the geometric approach generalizes to higher dimensions. The definition of a chain of horospheres extends in a straightforward fashion to N -dimensional hyperbolic space

$$\mathbb{H}^N = \{(x_1, x_2, \dots, x_N : x_N > 0),$$

and this gives us a concept of a continued fraction in higher dimensions. In contrast, the recurrence relations of (2.1) only make sense in the complex plane, and treating continued fractions algebraically in higher dimensions can be cumbersome.

REFERENCES

1. A. F. Beardon, *The Geometry of Discrete Groups*. Springer, New York, 1983, available at <http://link.springer.com/book/10.1007/2F978-1-4612-1146-4>.
2. ———, Continued fractions, discrete groups and complex dynamics, *Comput. Methods Funct. Theory* **1** (2001) 535–594, available at <http://link.springer.com/article/10.1007/2FBF03321006#>.
3. A. F. Beardon, M. Hockman, I. Short, Geodesic continued fractions, *Michigan Math. J.* **61** (2012) 133–150.
4. L. R. Ford, On the closeness of approach of complex rational fractions to a complex irrational number, *Trans. Amer. Math. Soc.* **27** (1925) 146–154.
5. ———, Fractions, *Amer. Math. Monthly* **45** (1938) 586–601.
6. G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*. Sixth edition. Oxford University Press, Oxford, 2008.
7. S. Katok, I. Ugarcovici, Symbolic dynamics for the modular surface and beyond, *Bull. Amer. Math. Soc.* **44** (2007) 87–132, available at <http://www.ams.org/journals/bull/2007-44-01/S0273-0979-06-01115-3/S0273-0979-06-01115-3/S0273-0979-06-01115-3.pdf>.
8. A. Ya. Khinchin, *Continued Fractions*. Translated from the third (1961) Russian edition. Reprint of the 1964 translation. Dover, Mineola, NY, 1997.
9. L. Lorentzen, H. Waadeland, *Continued Fractions*. Vol. 1. Second edition. Atlantis Studies in Mathematics for Engineering and Science, Atlantis Press, Paris, 2008.
10. C. Series, The modular surface and continued fractions, *J. London Math. Soc.* **31** no. 2 (1985) 69–80, available at <http://jllms.oxfordjournals.org/content/s2-31/1/69.extract>.

ALAN F. BEARDON received his Ph.D. from Imperial College, London, in 1964 and has taught at the University of Maryland, the University of Canterbury, and from 1968 to retirement, at the University of Cambridge.

His research interests are geometry, geometric function theory, discrete groups, dynamical systems, and mathematical economics. He is the author of seven books, and a winner of the Lester R. Ford prize for an expository paper in this MONTHLY.

*Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB,
United Kingdom
a.f.beardon@dpmms.cam.ac.uk*

IAN SHORT received his Ph.D. from the University of Cambridge in 2005, under the supervision of Alan Beardon. He is now a lecturer in mathematics at the Open University. His research interests include complex analysis, continued fractions, dynamical systems, and hyperbolic geometry.

*Department of Mathematics and Statistics, The Open University, Milton Keynes MK7 6AA, United Kingdom
ian.short@open.ac.uk*



Congratulations to MONTHLY Associate Editor, Edward B. Burger, who was inaugurated as the 15th President of Southwestern University on March 25, 2014.